arXiv:astro-ph/0402296v1 12 Feb 2004

THE COSMIC MICROWAVE BACKGROUND : PRESENT STATUS AND COSMOLOGICAL PERSPECTIVES

Cosmological Parameter Estimation : Method

Marian Douspis [a]

[a] **Astrophysics, Oxford University, Keble Road, OX1 3RH, Oxford, UK**
E-mail : douspis@astro.ox.ac.uk

Abstract.   CMB anisotropy data could put powerful constraints on theories of the evolution of our Universe. Using the observations of the large number of CMB experiments, many studies have put constraints on cosmological parameters assuming different frameworks. Assuming for example inflationary paradigm, one can compute the confidence intervals on the different components of the energy densities, or the age of the Universe, inferred by the current set of CMB observations. The aim of this note is to present some of the available methods to derive the cosmological parameters with their confidence intervals from the CMB data, as well as some practical issues to investigate large number of parameters. © 2003 Académie des sciences

keyword 1 / keyword 2 / etc.

Abstract.   Les observations des anisotropies du fond diffus cosmologique (FDC) peuvent placer de fortes contraintes sur les théories d'évolution de notre Univers. L'utilisation de telles données a permis de contraindre differents paramètres de différents cadres théoriques : l'age de l'U-nivers, son contenu baryonique, etc. Le but de cette contribution est de présenter differentes méthodes possibles pour extraire les paramètres cosmologiques et leurs intervalles de con-fiance des données du FDC. Des questions pratiques sur l'utilisation de grands nombres de paramètres sont aussi abordées. © 2003 Académie des sciences

keyword 1 / keyword 2 / etc.

1. Introduction

The extraction of information from cosmic microwave background (CMB) anisotropies is a classic problem of model testing and parameter estimation, the goals being to constrain the parameters of an assumed model and to decide if the *best–fit model* (parameter values) is indeed a good description of the data. Maximum likelihood is often used as the method of parameter estimations. Within the context of the class of models to be examined, the probability distribution of the data is maximized as a function of the model parameters, given the actual, observed data set[1]. Once found, the best model must then be judged on its ability to account for the data, which requires the construction of a *statistic* quantifying the *goodness–of–fit* (GoF). Finally, if the model is retained as a good fit, one defines *confidence* intervals on the parameter estimation. The exact meaning of these confidence intervals depends heavily on the method used to construct them, but the desire is always the same – one wishes to quantify the 'ability' of other parameters to explain the data (or not) as well as the best fit values. Given the quality of the current data, and the aim of the analysis – precise determination of the cosmological parameters – much attention should be put on the robustness and accuracy (unbiased techniques) of the methods used. I review the different ways of estimating the likelihood function of the parameters focusing on the use of the angular power spectrum (C_ℓ's). Then, some methods to compute the goodness of fit and the confidence intervals will be discussed. Finally, some practical issues for such computations will be addressed. In this review, I take the temperature fluctuations as the observed quantity. The same approaches could be applied for the polarisation signal of the CMB.

2. Likelihood

Data on the CMB consists of sky brightness measurements, usually given in terms of equivalent temperature in pixels. The likelihood function is to be constructed using these pixel values[2]. Standard Inflationary scenarios predict *Gaussian* sky fluctuations, which implies that the pixels should be modeled as random variables following a multivariate normal distribution, with covariance matrix given as a function of the model parameters (in addition to a noise term). It is important to note that, since the parameters enter through the covariance matrix in a non–linear way,the likelihood function $\mathcal{L}$ is not a linear function of the (cosmological) parameters.

Although it would seem straightforward to estimate model parameters directly with the likelihood function from the maps (*full analysis*), in practice the procedure is considerably complicated by the complexity of the model calculations and by the size of the data sets ([1, 2, 3, 4]). Maps consisting of several hundreds of thousands of pixels (the present situation) are extremely cumbersome to manipulate, and the million–pixel maps expected from Planck cannot be analyzed by this method in any practical way. An alternative is to first estimate the angular power spectrum from the pixel data and then work with this reduced set of numbers. For Gaussian fluctuations, there is in principle no loss of information. Because of the large reduction of the data ensemble to be manipulated, the tactic has been referred to as "radical compression" ([2]). The power spectrum has in fact become the standard way of reporting CMB results; it is the best visual way to understand the data, and in any case it is what is actually calculated in the models. The first part of this section describes briefly the full analysis procedure. Then, the second part will focus on the power spectrum as starting point for cosmological parameters estimation. In the latter case, due to the non-Gaussian behavior of the C_ℓ's, elaborated approximations should be used.

2.1. Full analysis

Temperature fluctuations of the CMB are described by a random field in two dimensions: $\Delta(\hat{n}) \equiv (\delta T/T)(\hat{n})$, where T refers to the temperature of the background and $\hat{n}$ is a unit vector on the sphere. It is

---

[1]In recent Bayesian analyses, quoting the mean of the product of the likelihood and prior functions as best model, is preferred
[2]The term pixel will be understood to also include temperature differences.

usual to expand this field using spherical harmonics:

$$\Delta(\hat{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n}) \qquad < a_{\ell m} a_{\ell' m'}^* >_{ens} = C_\ell \delta_{\ell\ell'} \delta_{mm'} \tag{1}$$

The $a_{\ell m}$'s are randomly selected from the probability distribution characterizing the process generating the perturbations. In the Inflation framework, which we consider here, the $a_{\ell m}$'s are *Gaussian random variables* with zero mean and covariance[3] given in Eq. 1. The C_ℓ's then represent the *angular power spectrum.* We may express the observed (or beam smeared) correlation between two points separated on the sky by an angle θ as

$$C_b(\theta) \equiv < \Delta_b(\hat{n}_1)\Delta_b(\hat{n}_2) >_{ens} = \frac{1}{4\pi} \sum_\ell (2\ell + 1) C_\ell B_\ell^2 P_\ell(\mu) \tag{2}$$

where P_ℓ is the Legendre polynomial of order l, $\mu = \cos\theta = \hat{n}_1 \cdot \hat{n}_2$ and B_ℓ is the harmonic coefficient of the beam decomposition[4]. The statistical isotropy of the perturbations demands that the correlation function depend only on separation, θ, which is in fact what permits such an expansion.

Given these relations and a CMB map, it is now straightforward to construct the likelihood function, whose role is to relate the N_{pix} observed sky temperatures, which we arrange in a *data vector* with elements $d_i \equiv \Delta_b(\hat{n}_i)$, to the model parameters, represented by a *parameter vector* $\overrightarrow{\Theta}$. For *Gaussian* fluctuations (with Gaussian noise) this is simply a multivariate Gaussian:

$$\mathcal{L}(\overrightarrow{\Theta}) \equiv \text{Prob}(\overrightarrow{d}\,|\,\overrightarrow{\Theta}) = \frac{1}{(2\pi)^{N_{\text{pix}}/2}|\boldsymbol{C}|^{1/2}} e^{-\frac{1}{2}\overrightarrow{d}^t \cdot \boldsymbol{C}^{-1} \cdot \overrightarrow{d}} \tag{3}$$

The first equality reminds us that the likelihood function is the probability of obtaining the data vector given the model as defined by its set of parameters. In this expression, $\boldsymbol{C}$ is the pixel covariance matrix:

$$C_{ij} \equiv < d_i d_j >_{ens} = T_{ij} + N_{ij} \tag{4}$$

where the expectation value is understood to be over the theoretical ensemble of all possible universes realizable with the same parameter vector. The second equality separates the model's pixel covariance, $\boldsymbol{T}$, from the noise induced covariance, $\boldsymbol{N}$. According to Eq. 2, $T_{ij} = C_b(\theta_{ij}) \equiv 1/(4\pi) \sum_\ell (2\ell + 1) C_\ell W_{ij}(\ell)$ where $\boldsymbol{W}$, the window matrix, contains the beam and strategy effects (direct measure, differences). The parameters may be either the individual C_ℓ (or band–powers, discussed below), or the fundamental cosmological constants, Ω, H_o, etc... In the latter situation, the parameter dependence enters through detailed relations of the kind $C_\ell[\overrightarrow{\Theta}]$, specified by the adopted model (e.g., Inflation).

For cosmological parameters estimations, one has to compute the likelihood value of Eq. 3 for a family of models investigated. For each set of parameters $\overrightarrow{\Theta}$, the computational time for the likelihood goes like N_{pix}^3, unless geometrical symmetries in the observational strategy allows to use faster algorithm for inverting $\boldsymbol{C}$. Investigating one handful of parameters with reasonable steps and ranges (typically $N_{parameters}^{10}$) with a map of few thousands of pixels becomes extremely cumbersome. Only few studies has been done in such a way ([16, 17, 18, 19]. Such a computation with second generation experiments is thus prohibitive[5].

2.2. Using the Angular Power Spectrum

In order to avoid the problem of computational cost of the full analysis, an alternative consists in first estimating the angular power spectrum from the pixel data and then work with the latter to estimate the cosmological parameters. The critical issue is then how to correctly use the power spectrum for an unbiased

---

[3]The indicated averages are to be taken over the theoretical ensemble of all possible anisotropy fields, of which our observed CMB sky is but one realization.

[4]Note that this expansion pre-supposes axial symmetry for the beam

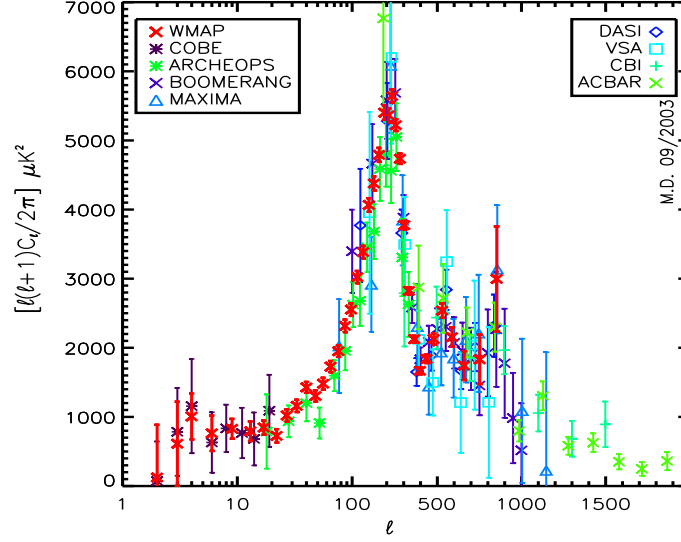[5]except for some particular symmetries, [5]

Figure 1: Angular power spectrum estimates of the CMB anisotropies in September 2003 ([7, 8, 9, 10, 11, 12, 13, 14, 15]). Notice the good agreement between band powers coming from different experiments (different detectors, technology, scanning strategy,...) until $\ell \sim 500$.

parameter estimation and model evaluation. The angular power spectrum can be evaluated with different techniques (see Hamilton, this issue, [6]). Again, a likelihood analysis from the maps can be done by inserting a spectral form into the definition of T. For example, the commonly used *flat* band–power, δT_{fb} (or $\mathcal{C}_b = \delta T_{fb}^2$ over a certain range in ℓ), actually represents the equivalent logarithmic power integrated over the band, which simplify the correlation matrix as follows:

$$C_\ell \equiv 2\pi[\delta T_{fb}^2/(\ell(\ell+1)] \qquad T = \frac{1}{2}\delta T_{\text{fb}}^2 \sum_\ell \frac{(2\ell+1)}{\ell(\ell+1)}W(\ell) \qquad (5)$$

In this way, we may write Eq. 3 in terms of the band–power and treat the latter as a parameter to be estimated. This then becomes the band–power likelihood function, $\mathcal{L}(\delta T_{\text{fb}})$. Figure 1 shows the latest band power estimates of the CMB fluctuations. Some of the points have been obtained by maximizing this likelihood function; the errors are typically found by in a Bayesian approach, by integration in $\mathcal{C}_b$ over $\mathcal{L}$ with a uniform prior (eg. DASI[12], VSA[13], CBI[14], ACBAR[15]). Other band powers and errors are estimated by using Monte Carlo based methods (see [26, 27]) like the WMAP[7], BOOMERANG[10], MAXIMA[11] and Archeops[9] ones. Notice that the variance due to the finite sample size (i.e., the sample variance, including the cosmic variance due to our observation of one realization of the sky) is fully incorporated into the analyses.

Given a set of band–powers how should one proceed to constrain the fundamental cosmological parameters, denoted in this subsection by $\overrightarrow{\Theta}$? If we had an expression for $\mathcal{L}(\overrightarrow{\delta T_{\text{fb}}})$, for our set of band–powers $\overrightarrow{\delta T_{\text{fb}}}$, then we could write $\mathcal{L}(\overrightarrow{\delta T_{\text{fb}}}) = Prob(\overrightarrow{d}|\overrightarrow{\delta T_{\text{fb}}}) = Prob(\overrightarrow{d}|\overrightarrow{\delta T_{\text{fb}}}[\overrightarrow{\Theta}]) = \mathcal{L}(\overrightarrow{\Theta})$. Thus, our problem is reduced to finding an expression for $\mathcal{L}(\overrightarrow{\delta T_{\text{fb}}})$, but as we have seen, this is a complicated function of $\overrightarrow{\delta T_{\text{fb}}}$, requiring use of all the measured pixel values and the full covariance matrix with noise – the very thing we are trying to avoid. Our task then is to find an approximation for $\mathcal{L}(\overrightarrow{\delta T_{\text{fb}}})$.

4

2.2..1 χ^2 minimization

The most obvious way of finding "the best model" given a set of points and errors is the traditional χ^2– minimization. This means that we assume a Gaussian shape for the likelihood function of the kind:

$$\mathcal{L}(\overrightarrow{\delta T_{\text{fb}}}) = e^{-\chi^2(\overrightarrow{\Theta})/2.} = e^{-(\overrightarrow{d}^{\,obs} - \overrightarrow{d}^{\,model}) \cdot \boldsymbol{M}^{-1} \cdot (\overrightarrow{d}^{\,obs} - \overrightarrow{d}^{\,model})/2.} \tag{6}$$

where $\boldsymbol{M}$ is the correlation matrix between the different flat band estimates. The main problem with this approach is that it deals with the flat band–power estimates as Gaussian distributed data which they are not (obeying the statistics of the square of a Gaussian). Then, it has been shown that such a procedure gives a biased estimation of the cosmological parameters and bad estimates of the confidence intervals ([2, 20]), leading to the search of more accurate approximation of the likelihood function.

2.2..2 More elaborated approximations

Different studies have been made to reconstruct better analytical approximations directly from the form of the flat band likelihood function ([2, 20, 21]). This section will focus on two of them. One is derived from the likelihood function in a particular case, for which it is actually exact [20]. The other one, mostly used during the last years, offers the advantage of being really close to a χ^2 minimization by changing variables in the appropriate way [2]. Both approximations need a small amount of information and aim to be used directly from the spectrum given in the literature[6].

- BDBL approximation

The Bartlett, Douspis, Blanchard and Le Dour approximation is based on the analytical form of the likelihood in an "ideal" experiment, where all the pixels (N_{pix}) are independent random variables (uncorrelated) and the noise is uncorrelated and uniform (σ_N^2). In that particular case, one can write the exact likelihood function as follows:

$$\mathcal{L}(\delta T_{\text{fb}}) \propto X^{\nu/2} e^{-X/2} \quad X[\delta T_{\text{fb}}] \equiv \frac{([\delta T_{\text{fb}}^{(o)}]^2 + \beta^2)}{([\delta T_{\text{fb}}]^2 + \beta^2)} \nu \tag{7}$$

where $\beta = \sigma_N$ and $\nu = N_{\text{pix}}$. The approximation comes from the fact that the authors keep the same likelihood form for real experiments, whereas the noise is no longer uniform and uncorrelated, and the pixels are not independent. To take into account these differences, one lets ν and β as free parameters and fixes them by fitting the 68% and 95% confidence intervals (published or inferred from the true likelihood function). Figure 2 shows the comparison between this approximation and true likelihood functions obtained for TOCO data ([23, 24]).

The advantage of this approximation is that the better the behavior of the experiment is (less correlations, more uniform noise, ...), the better the approximation is; being exact for the ideal case. It is also unbiased at the maximum of the likelihood and allows one to recover the full shape of the likelihood function. The inconvenient of this approximation is that it is defined only for uncorrelated flat band powers. The possible correlations between bands are not taken into account, as the full likelihood is given by the product of all individual likelihood functions: $\mathcal{L}(\overrightarrow{\delta T_{\text{fb}}}) = \prod \mathcal{L}(\delta T_{\text{fb}}^i)$.

- BJK approximation

In the second case, also referred as the Bond, Jaffe & Knox approximation the motivation is driven by the need to work with Gaussian distributed variables for which the χ^2 is not biased any more. Writing

---

[6]As we will see these approximations need one more information than the basic χ^2 minimization in order to take into account the non Gaussian behavior of the likelihood. The authors have been asking that this information is provided in addition to the band powers estimates and errors. Recent experiments have published the necessary information in their papers.
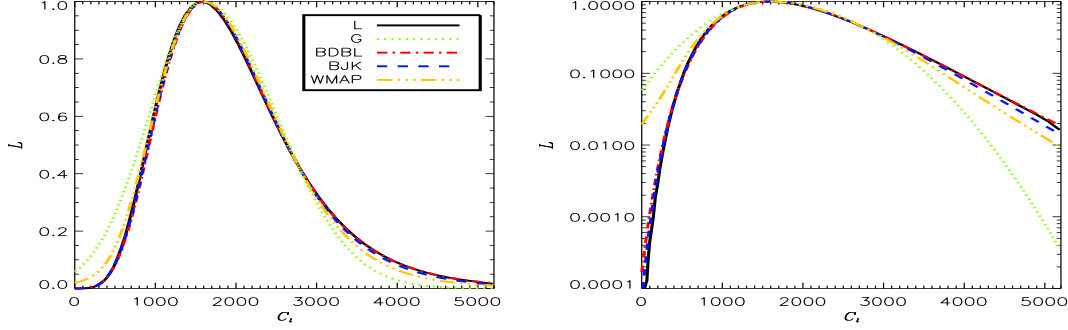
Figure 2: Comparison to the TOCO97 likelihood function for all approximation described in this section. The black solid line shows the full likelihood function computed from the map. The red dot–dashed and blue dashed lines show the BDBL and BJK approximation respectively. The 2–wings Gaussian and WMAP–type approximation are plotted in dotted green and dot-dot-dashed yellow lines

the likelihood in the spherical harmonic space for the same ideal experiment as above, and considering $\mathcal{C}_\ell = \ell(\ell+1)C_\ell/(2\pi) = \delta T_\ell^2$, $\mathcal{N}_\ell = \ell(\ell+1)N_\ell/(2\pi)$ where $N_\ell = \langle |n_{\ell m}|^2 \rangle$ is the noise power spectrum in spherical harmonics, one can show that the curvature matrix evaluated at the maximum is proportional to $\left(\mathcal{C}_\ell + \mathcal{N}_\ell/B_\ell^2 \right)^{-2} \delta_{\ell\ell'}$. If one define $Z_\ell \equiv \ln(\mathcal{C}_\ell + x_\ell)$ where in this particular "ideal experiment" $x_\ell = \mathcal{N}_\ell/B_\ell^2$, the curvature matrix expressed in term of Z is then constant. BJK approximation to the likelihood is then to take Z_b (determined in a band) as normally distributed in realistic experiments (by finding the good expression of the corresponding x_b). From the previous statements, one can then express the likelihood by:

$$\mathcal{L}(\overrightarrow{\delta T_{\rm fb}}) = e^{(-Z \cdot M^{-1} \cdot Z^t)/2.} \quad \text{where} \quad Z_i = \ln(\delta T_{\rm fb}^2(i) + x_b(i)) \quad \text{evaluated in a band } i \qquad (8)$$

The absolute value of Eq. 8 gives also an estimate of the goodness of fit. As we will see below, this approximation is slightly biased at the maximum of the likelihood but has been shown to be a reliable approximation (see Fig. 2) and is available online through the RADPACK package of .;[22].

The WMAP team [25] adopted an hybrid approximation: $\ln \mathcal{L} = \frac{1}{3} \ln \mathcal{L}_{Gauss} + \frac{2}{3} \ln \mathcal{L}_{BJK}$ motivated by an expansion of the true likelihood around the maximum[7]. This formulation has the advantage to be unbiased around the maximum but has not been tested against the real likelihood function in the wings.

Once the likelihood function $\mathcal{L}(\overrightarrow{\delta T_{\rm fb}})$ is known, one is able to compute $\mathcal{L}(\Theta)$ for a family of models. As we have seen, temperature on the sky are random Gaussian variables and then the "radical compression" is thus valid and induces no loss of information. The latter is true only if all the spectrum (in the limit of sensitivity of the experiment; the window function) is specified. Whereas, for different reasons (partial sky coverage, noise correlation, ...) only the spectrum *in band* is recovered: the spectrum is approximated by steps in ℓ. Such description induces a loss of information which may have some effect on the cosmological parameter estimation (bias and degeneracies). Douspis et al. ([19]) have shown that a better description of the spectrum (power in band *and* slope in band) could decrease the bias. The second and third generation of experiments provides (or will provide) better sensitivity, less correlated measurements which allows one to recover the spectrum with better resolution in ℓ (see WMAP for ex.), decreasing therefore the bias. Most of the studies are nevertheless performed by using the set of flat band estimates likelihood functions.

---

[7]where $\mathcal{L}_{Gauss} = \exp(-\chi^2/2)$

3. Goodness of fit

Once the (approximated) likelihood values of the models investigated are computed, one should find the best model (maximum) and evaluate the quality of the fit before constructing the parameter constraints. As a general rule, one must judge the quality of the fit before any serious consideration of the confidence intervals on parameters. This requires the application of a *Goodness of fit* (GOF) statistic. The latter is usually a function of both the model and the data, which reaches a maximum (or minimum) when the data is generated from the theory. The 'significance' may then be defined as the probability of obtaining $gof > gof_{\mathrm{obs}}$. On this basis, it permits a quantitative evaluation of the quality of the best model's fit to the data: if the probability of obtaining the observed value of the GOF statistic (from the actual data set) is low (low significance), then the model should be rejected. Without such a statistic, one does not know if the best model is a good model, or simply the "least bad" of the family.

In the full likelihood analysis method, the best model (set of parameters) could be obtained by maximizing the likelihood function of Eq 3 and is defined by $\overrightarrow{\Theta}_{best}$ in the following. One can easily note the Gaussian form of Eq 3 in the data vector $\overrightarrow{d}$. Given the best model, the most obvious GOF statistic is then clearly $gof = \overrightarrow{d}^{\,t} \cdot \tilde{\boldsymbol{C}}^{-1} \cdot \overrightarrow{d}$ where $\tilde{\boldsymbol{C}} \equiv C(\overrightarrow{\Theta}_{best})$ is the correlation matrix evaluated at the best model. For the Gaussian fluctuations we have assumed, this quantity follows a χ^2 distribution, with a number of degrees–of–freedom (DOF) approximately equal to the number of pixels minus the number of parameters[8].

The use of χ^2 method (in δT_{fb} or any change of variables like in BJK approximation) makes even easier the computation of the GOF. The obvious GOF statistics would just be one number, the value of the χ^2 evaluated at the minimum: $gof = \chi^2(\overrightarrow{\Theta}_{best})$. It is of course true that if the number of contributing effective DOF is large, a power estimate will closely follow a Gaussian; this, however, is never the case on the largest scales probed by a survey. Douspis et al., [28], have shown for example that the χ^2 approach leads to quantitatively different results than other, more appropriate GOF statistics.

When more elaborated approximations are used, the goodness of fit computation is less obvious. One should first reconstruct the distribution of the estimators. This could be a natural output when Monte Carlo based methods are used for the C_ℓ's extraction [26, 27], but it is mostly unknown when one applies traditional methods. Douspis et al. [28] have proposed an approximation which allows to reconstruct the distribution from the shape of the flat–band likelihood function[9]. When the latter is known, one should build a GOF statistique in order to compute to data probability given the best model (see [28] for examples).

Knowing that the best model is indeed a good fit to the data, or that the data have a good chance to be generated from this model, one should proceed by estimating the confidence intervals on the investigated parameters.

4. Confidence Intervals

The estimation of confidence intervals is mostly a question of definition. Most of CMB analyses have been done in the Bayesian framework and are thus dependent on the priors assumed. Some frequentist attempts have been performed in order to eliminate such dependencies. The reader can read more about the comparison between the two methods in [29].

Typically, the frequentist analyses are related to the goodness-of–fit statistics and the probability distribution of data for a given model (set of parameters). In the Bayesian approach, one reconstructs the conditional (posterior) probability density function (pdf), $P(\overrightarrow{\Theta}_{true}|d_{obs})$, for the unknown $\overrightarrow{\Theta}_{true}$ given the observation d_{obs}[10], from the pdf (which dependency in $\overrightarrow{\Theta}$ is known) for observing d using Bayes theorem. The latter

---

[8]This recipe does not strictly apply in the present case, because the parameters are non–linear functions of the data; it is nevertheless standard practice. In any case, the number of pixels is in practice much larger than the number of parameters. The numbers of degrees of freedom is also less than the number of pixels because of the correlations between pixels (non diagonal correlation matrix). Nevertheless, the matrices are mainly diagonal and the gof is then mostly insensitive to the small reduction of the number of DOF.

[9]This technique could be used both ways, allowing to reconstruct the likelihood when the distribution is known

[10]In our problem, d_{obs} should be taken as a set of flat band powers

evaluated at d_{obs} is known as Likelihood function: $P(d_{obs}|\overrightarrow{\Theta}) = \mathcal{L}(d_{obs}|\overrightarrow{\Theta})$.

$$P(\overrightarrow{\Theta}_{true}|d_{obs}) = \mathcal{L}(d_{obs}|\overrightarrow{\Theta}_{true})P(\overrightarrow{\Theta}_{true})/P(d_{obs}) \tag{9}$$

The denominator is just a normalization factor, and thus one of the issues is what to use for the prior $P(\overrightarrow{\Theta}_{true})$. If one knows the likelihood and fixes the prior (usually taken as uniform in terms of the parameters) then one knows the posterior probability distribution. A Bayesian credible region (interval) for a parameter is the range of parameter values that encloses a fixed amount of such probability. As the questions asked in the two approaches are quite different, one does not expect necessarily the intervals computed in the two methods to be similar.

I will described in the following two ways of estimating the confidence intervals referred as marginalisation and maximization. These two approaches are usually presented in opposition. In the limit of a Gaussian shaped likelihood with linear dependencies in the parameters, the two techniques are equivalent (see demonstration in [33]). Unfortunately this is not the case in cosmological parameter estimation. Both techniques consist in two steps: first one reduces the number of parameters in order to visualize the likelihood (or pdf) function (or surface in 2 dimensions). Then one computes the confidence intervals for each parameter.

4.1. Marginalisation

In CMB analyses, as $\overrightarrow{\Theta}$ is usually a vector of 5 to 10 parameters, it is quite hard to visualize the posterior (or likelihood) distribution. It is common then to retrieve one-dimensional probability by using an integration method (marginalisation).This technique is mostly used in Bayesian approaches to parameter estimation. Let's assume that $\overrightarrow{\Theta} = (x, y, ..., z)$ and we are interested in plotting the likelihood and finding the 68% confidence intervals on x, where the other parameters have been marginalised over. One usually computes:

$$\mathcal{L}(x) = \int ... \int \mathcal{L}(d_{obs}|(x, y, ..., z))P(x, y, ..., z)dyd...dz \tag{10}$$

$$\int_0^{x_m} \mathcal{L}(d_{obs}|x)dx = 0.5 \qquad \int_0^{x-} \mathcal{L}(d_{obs}|x)dx = 0.16 \quad \int_0^{x+} \mathcal{L}(d_{obs}|x)dx = 0.84 \tag{11}$$

where we assume that x is a positive variable, $\mathcal{L}$ is normalized to unity and P is a uniform prior on the parameters[11]. $[x-, x+]$ is then referred as the 68% confidence interval on x with all the other parameters marginalised over and x_m is quoted as the mean value (such computation of intervals in referred as EQT for "equi–probability tail")[12]. This may be seem easy in one dimension but could become cumbersome when dealing with 10 dimensional likelihood function (especially for the multi dimensional integral of the marginalisation Eq. 10). In order to be less and less dependent of all these effects, and to decrease the computational time of this step, maximization technique is mostly used.

4.2. Maximization

For the maximization technique, one assumes also a uniform prior in terms of the parameters (typically $P(x, y, ..., z) = 1$) but defines the pdf ($\equiv$ likelihood then) in one dimension as:

$$\mathcal{L}(x) = max_{y,...,z}[\mathcal{L}(d_{obs}|(x, y, ..., z))]P(x, y, ..., z) \tag{12}$$

---

[11]The prior is usually taken as uniform in $\overrightarrow{\Theta}$ in order to show our ignorance on the true value of $\overrightarrow{\Theta}$, even if there is no basis in Bayesian theory. In that sense, the interval will depend on the choice of parameters. Assuming $\overrightarrow{\gamma} = \overrightarrow{\Theta}^2$ as parameters and thus a uniform prior in $\overrightarrow{\gamma}$ will resume in a different interval (see Fig. 4.2.)

[12]Due to the non-linear dependency of the likelihood against the parameters, the shape of the latter could be highly non Gaussian. In such cases, it could occur that the maximum of the likelihood (described earlier as the best model) does not fall inside the 68% confidence interval (see for example Fig. 4.2.). In that case one should recompute the interval following the HPD (for "higher posterior distribution") technique, by fixing $\mathcal{L}(x-) = \mathcal{L}(x+)$ and $\int_{x-}^{x+} \mathcal{L}(d_{obs}|x)dx = 0.68$.
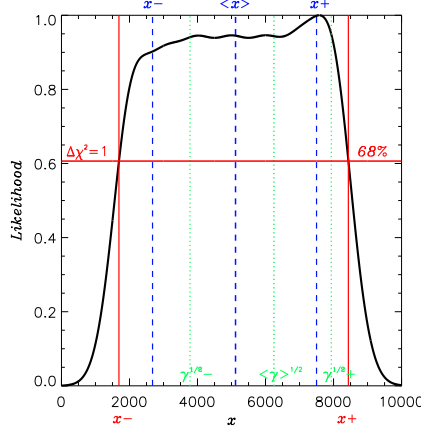
Figure 3: Comparison between marginalisation and maximization estimation of confidence intervals in an extreme case. The black solid curve is a one-dimensional likelihood function. The blue vertical dashed lines mark the mean and boundaries of 68% CL interval computed by integration ($[x-, x+]$). In that case the maximum of the likelihood $\hat{x} = 7580$ is just outside the interval. The green dotted lines are obtained with the same method but with taking $\gamma = x^2$ as variable (see text). Finally the red solid vertical lines shows the interval computed by taking values of the likelihood higher than $exp(-\Delta\chi^2/2) \times \mathcal{L}_{max}$, where here, $\mathcal{L}_{max} = 1$ and $\Delta\chi^2 = 1$.

which means that for each value of x one takes the maximum of the likelihood on all the other dimensions. Then, instead of integrating the resulting one dimensional likelihood like in Eq. 10 for obtaining the confidence intervals, one considers the values of the likelihood. For example, the boundaries of the 68% CL region are that where the likelihood has fallen by a factor $e^{-1/2}$ from its maximum, $\mathcal{L}_{max}$. As demonstrated in [33] this approximation becomes exact for multivariate Gaussian forms. One can define different CL intervals by choosing $\Delta\chi^2$ such as $\mathcal{L}(x_\alpha)/\mathcal{L}_{max} = e^{-\Delta_\alpha\chi^2/2}$ where α marks the confidence level. In one dimension, $\Delta_\alpha\chi^2 = 1, 4, 9$ for respectively $\alpha = 68, 95, 99\%$ CL. Fig. 4.2. shows an example in such a case. This technique does not give the real 68, 95 etc. confidence intervals, obtained only with Monte Carlo simulations by definition, but it is independent of the choice of the parameter (x versus x^2); the maximum of the likelihood is always inside every interval by definition, and it is computationally not consuming. Arguments and discussion about the different techniques can be found in [30].

5. Practical Issues

We have seen in the previous sections some of the existing statistical tools needed to perform a proper cosmological parameter estimation. As one would like to investigate a large number of parameters, and so a large number of models, some practical issues may be taken in consideration. I will describe in the following two methods (and some techniques) which correspond to the two actual ways of determining cosmological parameters from CMB anisotropies.

5.1. Cl's computations

The release of CMBFAST ([37]) has brought a major improvement in cosmological parameter estimations. The ability to compute a theoretical power spectrum in less than one minute (instead of one hour precedently) has allowed different groups to investigate many parameters in the same analysis. Different versions of the code have improved the first release by taking into account many physical effects (neutrino, reionisation, isocurvature modes, ...) and improved the computation by separating small scales effects from

large scales ones ("k-splitting"). Derived from this initial code, CAMB ([38]) increased the speed of computation by using FORTRAN 90 facilities. Finally, DASH ([39]) allows to compute C_ℓ's spectra in few seconds, by interpolating a precomputed grid of spectra in Fourier space. All these codes are more and more efficient and fast, and being adapted to be used in parallel computing.

5.2. Gridding

The gridding method consists in computing the likelihood values of different models following a periodical increment for each selected parameter, resulting in a N_{param} dimensional matrix. Historically, the parameter estimation from CMB anisotropies started with small grids of models, typically 3 or 4 free parameters with around 10 values each, the other ones fixed to the supposed best value of the moment [31, 32]. Then the number of parameters increased with the increasing speed of computer processors and the development of faster codes to compute the C_ℓ's (eg. CMBFAST)

One of the advantages of gridding is that one can compute a grid of models, store it and then compute the likelihood with one's set of data. If new data come out, one has just to compute the likelihood part again.

As the number of models investigated increases the storage could become a problem [33]. Then, some compression techniques, in combination with approximated interpolations, could be applied in order to store the necessary information only. The C_ℓ's computation time may also become a problem. There again approximations based on the known behavior of the C_ℓ's with parameters have been developed [33].

One of the inconvenients of the gridding method is that the position of the maximum of the likelihood grid is highly dependent of the grid itself. Namely, the maximum falls necessarily on one point of the grid. This effect is also recurrent when one uses the maximization technique. In order to avoid this, spline interpolation techniques are used when looking for the maximum along one or more parameters [33].

Finally, by definition, the gridding method is well adapted to multi–processors and data–grid method.

5.3. Monte Carlo Markov Chains

During the last few years, as an alternative to the gridding method, the Markov Chain Monte Carlo (MCMC) likelihood analyses had become a powerful tool in cosmological parameter estimation. This method generates random draws from the posterior distribution that is supposed to be a "realistic" sample of the likelihood hypersurface. The mean, variance, confidence levels can then be derived from this sample. Unlike the gridding method, scaling exponentially with the number of parameters, the MCMC method scales linearly with N_{param} allowing one to explore a larger set of parameters or to do the analysis faster.

Two issues should be highlighted in this method. The first one is the step in the random sampling. Typically, the step is taken as the standard deviation for each parameter. If it is too large, the chain can take a infinite time to converge and the acceptance rate is very low. If it is too small the chain will be highly correlated leading also to a slow convergence. A second issue is the convergence of the chain. At the beginning the sampling of the likelihood is very correlated and is not a "fair" representation of the posterior distribution. After a "burning period", the chain converges, the samples are independent and the likelihood function could be retrieve. The criterium of convergence is not a well defined quantity.

More explanations and applications could be found in [34, 35] and a FORTRAN 90 set of routines is available online [36].

6. Conclusions

In order to derive the cosmological parameters in a given framework from the temperature fluctuation of the CMB, many steps are needed. When the observed power spectrum is derived, one could use different techniques to estimate successively the (approximated) likelihood value of the family of models (parameters) investigated, the best model and its goodness of fit, and finally the confidence intervals on each parameter. Each of these steps may be highly cpu and memory consuming. With better and better observations, sensitivity and sky coverage, brute force maximum likelihood methods become impossible. Many approximations and techniques have then been developed during the last years, allowing to analyze

more and more data with increasing speed. When the appropriate method is used, this leads to an unbiased estimate of the cosmological parameters. These developments have demonstrated that efficient methods could be developed to take full advantage of data at the Planck accuracy and allow to determine parameters of cosmological relevance to a remarkably high accuracy. This is opening the golden road of precision cosmology.

References

[1] Bond J.R., Jaffe A.H. & Knox L. 1998a, Ph.Rev D57, 2117

[2] Bond, J. R., Jaffe, A. H. & Knox, L. 2000, ApJ, 533, 19

[3] Borrill J. 1999b, in: Proceedings of the 5th European SGI/Cray MPP Workshop, astro–ph/9911389

[4] Kogut A. 1999, ApJ 520, L83

[5] Wandelt, B. D. & Hansen, F. K. 2003, *Phys. Rev. D*, 67, 23001

[6] Hamilton, CR Physique (2003), this issue

[7] Hinshaw, G. et al. 2003, *ApJS*, 148, 135

[8] Tegmark M. & Hamilton A. 1997, [`astro--ph/9702019`]

[9] Benoît, A. et al. 2003, *A&A*, 399, L19

[10] Ruhl, J. E. et al. (BOOMERanG collab.) 2002, [`astro-ph/0212229`]

[11] Hanany, S. et al. (MAXIMA collab.) 2000, *ApJ*, **545**, L5

[12] Halverson, N. W. et al. (DASI collab.) 2002, *ApJ*, **568**, 38

[13] Grainge, K. et al. (VSA collab.) 2003, [`astro-ph/0212495`]

[14] Pearson, T. J. et al. 2003, *ApJ*, 591, 556

[15] Kuo, C. L. et al. (ACBAR collab.) 2002, [`astro-ph/0212289`]

[16] Ratra, B. et al., 1999, *ApJ*, 517, 549

[17] Rocha, G. et al., 1999, *ApJ*, 525, 1

[18] Górski K.M. 1994, ApJ 430, L85

[19] Douspis M., Bartlett, J.G., Blanchard A. & Le Dour M., Astronomy and Astrophysics, v.368, p.1-14, 2001a

[20] Bartlett, J. G., Douspis, M., Blanchard, A., & Le Dour M. 2000, A&AS, 146, 507

[21] Wandelt, B. D., Hivon, E., & Górski, K. M. 2001, *Phys. Rev. D*, 64, 83003

[22] web: http://bubba.ucdavis.edu~knox/radpack.html

[23] Miller, A. D. et al. 1999, *ApJ*, 524, L1

[24] web: http://imogen.princeton.edu/mat/results/

[25] Verde et al. (WMAP collaboration) 2003, accepted in ApJ, astro–ph/0302218

[26] Hivon, E., Górski, K. M., Netterfield, C. B., Crill, B. P., Prunet, S., & Hansen, F. 2002, *ApJ*, 567, 2

[27] Szapudi, I., Prunet, S., Pogosyan, D., Szalay, A. S., & Bond, J. R. 2001, *ApJ*, 548, L115

[28] Douspis, M., Bartlett, J. G. & Blanchard, 2003, A&A in press, astro–ph/0305428

[29] Abroe, M. E. et al. 2002, *MNRAS*, 334, 11

[30] Feldman, G. J. & Cousins, R. D. 1998, *Phys. Rev. D*, 57, 3873

[31] Lineweaver C., Barbosa D., Blanchard A. & Bartlett J.G. 1997, A&A 322, 365

[32] Hancock S., Rocha G., Lasenby A.N. & Gutierrez C.M. 1998, MNRAS 294, L1

[33] Tegmark, M. & Zaldarriaga, M. 2000, *ApJ*, 544, 30

[34] Christensen, N. & Meyer, R. 2001, *Phys. Rev. D*, 64, 22001

[35] Lewis, A. & Bridle, S. 2002, *Phys. Rev. D*, 66, 103511

[36] web: http://cosmologist.info/cosmomc/

[37] Seljak U. & Zaldarriaga M. 1996, ApJ469, 437

[38] Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

[39] Kaplinghat, M., Knox, L., & Skordis, C. 2002, *ApJ*, 578, 665